Supplementary material

Thousands of primer-free, high-quality, full-length SSU rRNA sequences from all domains of life

Søren M. Karst, Morten S. Dueholm, Simon J. McIlroy, Rasmus H. Kirkegaard, Per H. Nielsen and Mads Albertsen

Center for Microbial Communities, Department of Chemistry and Bioscience, Aalborg University, Denmark

Materials and methods

A detailed overview of the workflow can be found **Fig. S1**.

General molecular methods:

RNA and DNA concentrations were determined using the Qubit® RNA or dsDNA BR or HS assay kits and a Qubit 2.0 fluorometer (Thermo Fisher Scientific). RNA quality was assessed using Tapestation 2200 gel electrophoresis and RNA screentapes (Agilent). RiboLock RNase inhibitor (Thermo Fisher Scientific) was added to the purified total RNA to minimize RNA degradation. All commercial kits were used according to the protocol provided by the manufacturer unless otherwise stated. Primers and adaptors used in this study can be found in **Table S5**.

Samples and nucleic acid purification:

Mock community *Escherichia coli* MG 1655 (DSM 18039) and *Bacillus subtilis* subsp. *subtilis* str. 168 (DSM 402) were obtained from the German collection of microorganisms and cell cultures (DSMZ). *P. aeruginosa* PAO1 was obtained from the Seattle *P. aeruginosa* PAO1 transposon mutant library[1]. The bacteria were streaked on LB agar plates, which were then incubated at 30°C (*B. subtilis*) or 37°C (*E. coli* and *P. aeruginosa*) until colonies appeared. A single colony of each bacterium was used to inoculate 50 mL of LB-media in a 250 mL conical flask, and the bacteria were grown overnight at the temperature provided above with 150 rpm of shaking. The cultures were divided into 1.5 mL aliquots and bacteria pelleted by centrifugation (16,200 x g, 4°C, 1 min). The cell pellets were stored at – 20°C. RNA was purified from the cell pellets using the RiboPure RNA purification kit for bacteria (Thermo Fisher Scientific). RNA concentration and quality were assessed as described above. Equal amounts for total RNA from each bacterium were mixed to create the mock community RNA sample.

Freshwater. The sample was collected from a pond at the Aalborg University campus (57.015436° N, 9.977698° E) in Denmark on 20 June 2016. The sample was prefiltered through a 5 µm polypropylene filter (Merck Millipore, Ireland) to remove algae and particular matters. Bacteria were then collected from 100 mL of the prefiltered water on 0.2 µm cellulose nitrate filters (Sartorius Stedim Biotech, Germany). RNA and DNA were isolated using the RNA PowerSoil® total RNA isolation kit with the RNA PowerSoil® DNA elution accessory kit (MO BIO, CA, USA) according to the manufacturer recommendations, except that the soil was replaced with two filters from above that had been rolled

into cylinders with the top side facing inward. RNA concentration and quality were assessed as described above.

Soil. The soil sample was collected at the Aalborg University campus (57.013684° N, 9.984724° E) in Denmark on 16 May 2016. RNA and DNA were immediately isolated using the RNA PowerSoil® total RNA isolation kit with the RNA PowerSoil® DNA elution accessory kit (MO BIO, CA, USA). RNA concentration and quality were assessed as described above.

Human gut sample. Eight grams of feces were obtained from a healthy male volunteer and immediately mixed with 16 mL of RNAlater (Thermo Fisher Scientific) by vigorous shaking in a 50 mL tube. The sample was centrifuged (700 x g, 4°C, 5 min) to remove solid fecal matters. Bacteria from 2 mL of the supernatant were pelleted by centrifugation (9,000 x g, 4°C, 5 min). Total nucleic acid was immediately extracted from the pellet using the PowerMicrobiome RNA isolation kit (MO BIO, CA, USA) with the optional phenol-based lysis. Purification was carried out according to the manufacturers' recommendations with the small modification that cell lysis was carried out in a FastPrep-24 instrument for 4x 40 s at 6.0 m/s to increase the yield of nucleic acids from bacteria with tough cell walls[2]. The sample was incubated on ice for 2 min between each bead beating to prevent heating of the sample. DNA-free total RNA was obtained by treating a subsample of the purified nucleic acid with the DNase Max kit (MO BIO, CA, USA). RNA concentration and quality were assessed as described above.

Anaerobic digestion. Anaerobic digestion biomass was sampled from an anaerobic digester at Viborg biogas plant at Viborg wastewater treatment plant (WWTP) (56.4251° N, 9.4543° E) in Denmark the 13th of September 2013. The plant treats primary sludge and surplus activated sludge from Viborg wastewater treatments plant under mesophilic conditions. The biomass was transported on ice and stored as 2 mL aliquots at – 80°C. Total nucleic acids were purified from 500 µL of sample using the PowerMicrobiome RNA isolation kit (MO BIO, CA, USA) as described for the human gut sample. DNA-free total RNA was obtained by treating a subsample of the purified nucleic acid with the DNase Max kit (MO BIO, CA, USA). RNA concentration and quality was assessed as described above.

Activated sludge. Activated sludge was sampled from an aeration tank at Aalborg West WWTP (57.049422° N, 9.864735° E) in Aalborg, Denmark on 31 October 2012. The plant primarily receives domestic wastewater and employ nitrogen and enhanced biological phosphorus removal. Sample material was transported on ice and processed within 4 h. The sample was homogenized (1 min, 1650 rpm) using a mortar-pestle, glass/teflon tissue homogenizer (30 mL) mounted on a Heidolph RZR 2020 (Heidolph, Germany) and stored as 2 mL aliquots at -80°C. Total nucleic acids were purified from 500 µL of sample using the PowerMicrobiome RNA isolation kit (MO BIO, CA, USA) as described for the human gut sample. DNA-free total RNA was obtained by treating a subsample of the purified nucleic acid with the DNase Max kit (MO BIO, CA, USA). RNA concentration and quality were assessed as described above.

cDNA synthesis

SSU rRNA size selection. SSU rRNA was enriched from total RNA by size selection on a E-gel electrophoresis system with precast E-Gel CloneWell gels (Thermo Fisher Scientific). Standard operation procedure was used with the following exceptions. Total RNA was heat denatured at 70°C for 5 min and snap cooled on ice block for 2 min to prevent RNA from refolding. Up to 800 ng total RNA sample was added per well. A total of 500 ng GeneRuler 1 kb DNA ladder was used as a reference. The gel was run until the SSU peak (ca. 1,500 bp) was 1 mm from the elution well (approx. 22 min run time), after which

elution aliquots were sampled every 15 seconds up to a total of 16 aliquots and the visible passing of the SSU rRNA peak. Every two aliquots were pooled (1+2, 3+4 etc.) and analyzed with Tapestation 2200 gel electrophoresis using High Sensitivity RNA Screentape (Agilent). Based on the gel electrophoresis results the aliquots containing SSU rRNA were pooled, and purified using Ampure XP beads (Beckman Coulter) and the protocol for PCR purification with the following exceptions. Sample to bead solution ratio was 1:1 and beads were washed with 80% EtOH. The purified SSU rRNA was eluted in 32 µL nuclease free water.

Poly(A) tailing of SSU rRNA. Poly(A) tail was added to the purified SSU rRNA using 30 µL purified SSU rRNA, 4 µL 10X *E. coli* Poly(A) Polymerase Reaction Buffer (NEB), 4 µL ATP (10mM) (NEB), 1.5 µL *E. coli* Poly(A) Polymerase (NEB) and 1 µL RiboLock RNase Inhibitor. The reaction was incubated at 37°C for 30 min. The product was immediately purified using Ampure XP beads. Sample to bead solution ratio was 1:1, and beads were washed with 80% EtOH. The purified poly(A) SSU rRNA was eluted in 40 µL nuclease free water. Samples were validated with Tapestation 2200 gel electrophoresis using RNA Screentape. Based on gel electrophoresis results samples were diluted to approx. 1 ng/µL using nuclease-free water.

First strand synthesis. First a priming adaptor (containing unique tag) was annealed to the poly(A) SSU rRNA in a reaction containing 11.5 µL poly(A) SSU rRNA, 0.5 µL SSU rRNA RT adaptor (100 µM), 1 µL 10 mM dNTP mix. The priming reaction was incubated at 70°C for 5 min and then snap-cooled for 2 min on an ice block. Annealing was followed by reverse transcription reaction, which contained 13 µL primed SSU rRNA samples, 4 µL 5x SSIV buffer, 1 µL SuperScript IV Reverse Transcriptase (200 U/µL) (Thermo Fisher Scientific), 1 µL 100 mM DTT, and 1 µL RiboLock RNase Inhibitor. The reverse transcription incubation was at 50°C for 50 min, inactivation at 80°C for 10 min, and hold at 37°C, at which point 1 µL Ribonuclease H, from *E. coli* (cloned) 10 U/µL (Thermo Fisher Scientific) was added, immediately followed by incubation at 37°C for 20 min to digest the RNA. The cDNA was purified using standard Ampure XP bead protocol with the following exceptions: Sample to bead solution ratio was 1:1, and beads were washed with 80% EtOH. The purified cDNA was eluted in 20 µL nuclease-free water.

Second strand synthesis. A priming site was added to the single-stranded cDNA by single-stranded ligation of an oligo to the cDNA. The single-stranded ligation reaction contained 8.25 µL eluted cDNA, 2.5 µL SSU_rRNA_s oligo (10 µM), 5 µL 10x Custom T4 RNA ligase buffer[3] (500 mM Tris-HCl, 100 mM $MgCl_2$, 10 mM hexammine cobalt chloride, 200 µM ATP, pH 8.0) , 31.25 µL 40% (wt/vol) PEG 8000, 0.5 µL BSA (1 mg/mL), 2.5 µL T4 RNA ligase (10 U/µL), and incubation was performed at 22°C for approximately 14 hrs. The cDNA was purified using standard Ampure XP bead protocol with the following exceptions. Sample to bead solution ratio was 1:1 and beads were washed with 80% EtOH. The purified product was eluted in 12 µL nuclease free water. The single-stranded ligated cDNA was converted to dsDNA using DNA polymerase. The dsDNA reaction contained 33.75 µL nuclease free water, 5 µL 10X Coral PCR Buffer (Qiagen), 1 µL dNTP (10 mM), 2.5 µL SSU_rRNA_l (10 µM), 2.5 µL f16S_rRNA_PCR_fw2 (10 µM), 2 µL $MgCl_2$ (25 mM), 0.25 µL Taq polymerase (5 U/µL) (Qiagen) and 3 µL single-stranded ligated cDNA. The reaction was initiated by denaturation at 94°C for 3 min and 2 cycles of denaturation at 94°C for 30 seconds, primer annealing at 56°C for 30 seconds, and extension at 72°C for 3 min. The cDNA was purified using standard Ampure XP bead protocol with the following exceptions: Sample to bead solution ratio was 5:3, and beads were washed with 80% EtOH. The purified product was eluted in 25 µL nuclease free water. None converted ssDNA was removed by single-stranded nuclease digestion to improve the efficiency of the subsequent PCR reaction. The single-stranded nuclease digestions reaction contained 23 µL purified dsDNA product, 3 µL 10X S1 Nuclease

buffer (Thermo Fisher Scientific), 3 µL 3M NaCl, 1 µL S1 nuclease 10 U/µL (Invitrogen). The reaction was incubated at 25 °C for 25 min and terminated by adding 2 µL 0.5 M EDTA and heated to 70°C for 10 min. The digested product was diluted to 50 µL with nuclease free water and purified using standard Ampure XP bead protocol with the following exceptions: Sample to bead solution ratio was 5:3 and beads were washed with 80% EtOH. The purified product was eluted in 12 µL nuclease-free water.

Read-tag and linked-tag library preparation

PCR amplification 1. The ds cDNA was amplified using PCR to obtain enough product for validation and size selection. The reaction contained 63.5 µL nuclease-free water, 10 µL 10X Coral PCR buffer, 2 µL 10 mM dNTP, 5 µL f16S_rRNA_PCR_fw 10 µM, 5 µL f16S_rRNA_PCR_rv 10 µM, 4 µL MgCl 25 mM, 0.5 µL Taq polymerase 5 U/µL and 10 µL ds cDNA. The reaction was initiated by denaturation at 94°C for 3 min and 20 cycles of denaturation at 94°C for 30 seconds, primer annealing at 62°C for 30 seconds and extension at 72°C for 2 min, and finishing with final extension at 72°C for 5 min. The PCR product was purified using standard Ampure XP bead protocol with the following exceptions. Sample to bead solution ratio was 5:3, and beads were washed with 80% EtOH. The purified product was eluted in 20 µL nuclease-free water. The cDNA amplicon was validated using Tapestation 2200 gel electrophoresis using standard protocol for D5000 screentapes.

PCR product size selection. Presence of none-SSU products after the first PCR amplification was frequently observed. To increase the final yield of full-length SSU rRNA amplicons, a size selection on a E-gel electrophoresis system with precast E-Gel CloneWell gels was performed. Standard operation procedure was used with the following exceptions: up to 800 ng of sample was added per well. A total of 500 ng GeneRuler 1 kb DNA ladder was used as a reference. The gel was run until the SSU peak (ca. 1500 bp) was 1 mm from the elution well (approx. 22 min run time), after which elution aliquots were sampled every 15 seconds up to a total of 16 aliquots and the visible passing of the SSU rRNA amplicon peak. Every two aliquots were pooled (1+2, 3+4 etc.) and validated with Tapestation 2200 gel electrophoresis using High Sensitivity D5000 Screentapes (Agilent). Based on the gel electrophoresis results the aliquots containing full-length SSU rRNA amplicons were pooled, and purified using Ampure XP beads (Beckman Coulter) and the protocol for PCR purification with the following exceptions: Sample to bead solution ratio was 5:3, and beads were washed with 80% EtOH. The purified SSU rRNA was eluted in 15 µL nuclease free water. The size selected cDNA amplicon product was validated with Tapestation 2200 gel electrophoresis and High sensitivity D5000 screentapes, and DNA concentration was measured using Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific) on a Qubit Fluorometer 2.0. Based on the DNA concentration, the molecule concentration was calculated and used to dilute the cDNA amplicon to 2,000 molecules/µL with nuclease-free water.

PCR amplification 2. The cDNA amplicons were amplified using PCR to obtain > 10,000 clonal copies of each cDNA amplicon molecule. For samples analyzed on the MiSeq a total of 20,000 – 30,000 molecules were used as template for the PCR. For the soil sample analyzed on the HiSeq, 300,000 molecules were used as template. For the mock community sample analyzed on the MinION (Oxford Nanopore Technologies), 5,000 molecules were used as template. The PCR reaction contained 63.5 µL nuclease free water, 10 µL 10X Coral PCR buffer, 2 µL 10 mM dNTP, 5 µL SSU_rRNA_pcr_fw 10 µM, 5 µL SSU_rRNA_pcr_rv 10 µM, 4 µL MgCl 25 mM, 0.5 µL Taq polymerase 5 U/µL and 10 µL diluted cDNA amplicon. The reaction was initiated by denaturation at 94°C for 3 min and 25-35 cycles of denaturation at 94°C for 30 seconds, primer annealing at 62°C for 30 seconds and extension at 72°C for 2 min and

finishing with final extension at 72°C for 5 min. The PCR product was purified using standard Ampure XP bead protocol with the following exceptions: Sample to bead solution ratio was 5:3, and beads were washed with 80% EtOH. The cDNA amplicon product was eluted in 20 µL nuclease-free water. The product was validated using Tapestation 2200 gel electrophoresis using standard protocol for D5000 screentapes. DNA concentration was measured using Qubit dsDNA HS Assay Kit on a Qubit Fluorometer 2.0.

Read-tag library preparation. A Nextera library preparation kit (Illumina) was used to prepare a paired-end read-tag sequencing library from the PCR2 cDNA amplicon product using a customized protocol. The cDNA amplicon product was diluted to 4.35 ng/µL in 11.5 µL nuclease free water (total of 50 ng DNA). A tagmentation reaction was prepared containing 11.5 µL 4.35 ng/µL cDNA amplicon, 12.5 µL tagment DNA buffer and 1.5 µL tagment DNA enzyme. The reaction was incubated at 55°C for 5 min. Immediately after tagmentation, the product was diluted to 100 µL and purified using standard Ampure XP bead protocol with the following exceptions: Sample to bead solution ratio was 5:3, and beads were washed with 80% EtOH. The cDNA amplicon product was eluted in 22 µL nuclease free water. The tagmentation fragment products were PCR amplified using two separate PCR reactions performed simultaneously. PCR reaction A selectively amplified fragments with the 5' termini and reaction B selectively amplified fragments with the 3' termini. PCR reaction A contained 2.5 µL S504 nextera adaptor, 2.5 µL 10 µM SSU_rRNA_readtag_fw adaptor, 7.5 µL Nextera PCR master mix, 2.5 µL PCR primer cocktail, and 10 µL purified tagmentation product. PCR reaction B contained 2.5 µL S504 nextera adaptor, 2.5 µL 10 µM SSU_rRNA_readtag_rv adaptor, 7.5 µL Nextera PCR master mix, 2.5 µL PCR primer cocktail, and 10 µL purified tagmentation product. The following PCR program was used: Initial elongation at 72°C for 3 min, initial denaturation at 98°C for 30 seconds, and 8 cycles of denaturation at 98°C for 10 seconds, annealing at 60°C for 30 seconds and elongation at 72°C for 3 min. The PCR A and PCR B reactions were purified using standard Ampure XP bead protocol with the following exceptions: Sample to bead solution ratio was 5:3, and beads were washed with 80% EtOH. The read-tag library products were eluted in 100 µL nuclease free water. The purifications were repeated and the products each eluted in 18 µL nuclease free water. The read-tag A and B libraries were validated using Tapestation 2200 gel electrophoresis and standard protocol for D5000 screentapes. DNA concentration was measured using Qubit dsDNA HS Assay Kit on a Qubit Fluorometer 2.0. The read-tag A and B libraries were stored at –20°C until sequencing.

Linked-tag Library preparation. A linked-tag library was prepared using a custom protocol using cDNA amplicons from PCR2 as input. The cDNA amplicons from PCR2 were end-repaired in a reaction containing 20.25 µL nuclease-free water, 2.5 µL 10X NEBNext End repair Reaction Buffer (New England Biolabs), 1.25 µL NEBNext End Repair Enzyme Mix (New England Biolabs), 1 µL PCR2 cDNA amplicon (10 ng of DNA). The reaction was incubated at 20°C for 30 min. The end-repair reaction was purified using standard Ampure XP bead protocol with the following exceptions. Sample to bead solution ratio was 1:1, and beads were washed with 80% EtOH. The end repaired cDNA amplicon product was eluted in 10 µL nuclease-free water. The end-repaired cDNA was circularized in an intra molecular blunt end ligation reaction containing 75 µL nuclease free water, 10 µL 50% (w/v) PEG 4000 solution, 10 µL 10X T4 DNA ligase buffer (New England Biolabs), 4 µL T4 DNA ligase (New England Biolabs), and 1 µL of 1 ng/µL end-repaired cDNA amplicon. The reaction was incubated at 16°C for 60 min. The circularized cDNA amplicons were purified using standard Ampure XP bead protocol with the following exceptions: Sample to bead solution ratio was 1:1 and beads were washed with 80% EtOH. The circularized cDNA amplicons

were eluted in 20 µL nuclease free water. The circularization of the cDNA amplicons fuses the unique tag containing termini together. This enables PCR amplification of the junction sequence, which produces an amplicon containing both unique tags. The tag-junction PCR contained 30.25 µL nuclease free water, 5 µL 10x Coral PCR buffer, 1 µL 10 mM dNTP mix, 1.25 µL 10 µM SSU_rRNA_linktag_fw, 1.25 µL 10 µM SSU_rRNA_linktag_rv, 3 µL 25 mM $MgCl_2$, 0.25 µL 5u/µLTaq polymerase. The PCR reaction was initiated by denaturation at 94°C for 3 min and 20 cycles of denaturation at 94°C for 20 seconds, primer annealing at 56°C for 20 seconds, and extension at 72°C for 20 seconds and finishing with final extension at 72°C for 3 min. The PCR product was purified using standard Ampure XP bead protocol with the following exceptions. Sample to bead solution ratio was 1:1 and beads were washed with 80% EtOH. The resulting linked-tag library product was eluted in 15 µL nuclease free water. The linked-tag library was validated using Tapestation 2200 gel electrophoresis and standard protocol for D5000 screentapes. DNA concentration was measured using Qubit dsDNA HS Assay Kit on a Qubit Fluorometer 2.0. The linked-tag libraries were stored at –20°C until sequencing.

Illumina sequencing

Illumina MiSeq sequencing. The read-tag library A and B were diluted to 3.1 ng/µL. The linked-tag library was diluted to 0.6 ng/µL. The libraries were pooled by combining 4.6 µL read-tag library A, 4.6 µL read-tag library B, and 0.8 µL linked-tag library. The library pool were paired-end (240 bp x 25 bp) sequenced on a MiSeq instrument (Illumina) using a MiSeq® Reagent Kit v3, 150 cycles (Illumina). The MiSeq was running Control Software v2.5.0.5 (Illumina), MiSeq Reporter v2.5.0.5 (Illumina) and Real Time analysis v1.18.54 (Illumina). The libraries were prepared and loaded on the MiSeq using the standard procedure (MiSeq System Guide, Document # 15027617 v01; Preparing Libraries for Sequencing on the MiSeq, document # 15039740) with the following changes. 5 µL library pool was denatured by adding 5 µL 0.1 N NaOH solution, mixing well by pipetting, and incubating for 5 min at 25°C. The denatured library pool was diluted by adding 990 µL cold Hybridization Buffer (Illumina). 570 µL denatured and diluted library pool was mixed with 30 µL denatured and diluted 20 pM Phix Control v3 library (Illumina), and loaded on the MiSeq reagent cartridge sample loading well no. 17. Custom read2 primer mix was prepared by mixing 3.4 µL SSU_rRNA_read2_fw (100 µM) and 3.4 µL SSU_rRNA_read2_rv (100 µM), and diluting with 600 µL Hybridization Buffer. The custom read2 primer mix was transferred to the MiSeq reagent cartridge well no. 20. To increase the number of sequencing cycles from 150 to 265 cycles 3.5 mL incorporation mix was added to well nr. 1 and 3.6 mL of scan mix to well no. 2. Spare incorporation mix and scan mix can be obtained from another MiSeq Reagent Kit v3 (Illumina). A samplesheet was prepared by using the Illumina Experiment Manager 1.9 software (Illumina). After starting the software, the buttons were pressed in following the chronology “Create Sample Sheet” > “MiSeq” > “Next” > “Other” > “FASTQ Only”. At the “Sample Sheet Wizard – Workflow Parameters” screen under the “FASTQ Only Run Settings” area, the following were filled out. Reagent Cartridge Barcode* = Custom name usually containing date. Sample Prep Kit = Nextera. Index Reads = 0. Read Type = Paired End. Cycles Read 1 = 240. Cycles Read 2 = 25. The other fields were not mandatory. Only box that was ticked in the “FASTQ Only Workflow-Specific Settings” area was the “Custom Primer for Read 2” box. Afterwards “Next” was pressed. In the “Sample Sheet Wizard – Sample Selection” screen, the sample ID was typed in, in the field called “Sample ID*” in the “Samples to include in sample sheet” area. Afterwards “finish” was pressed, and the samplesheet was saved. When setting up the sequencing on the MiSeq, and the MiSeq warned that the MiSeq Reagent Kit v3 (150 cycles) is not compatible with 265 cycles, “Continue” was pressed, ignoring the warning, since extra reagent for more cycles had been

added. After sequencing, bcl2fastq v2.17.1.14 (Illumina) was used to generate FASTQ files from bcl files using standard settings (bcl2fastq Conversion User Guide, Part # 15038058 RevB).

Illumina HiSeq sequencing. The read-tag library A and B were diluted to 3.1 ng/µL. The linked-tag library was diluted to 0.6 ng/µL. The libraries were pooled by combining 4.6 µL read-tag library A, 4.6 µL read-tag library B, and 0.8 µL linked-tag library. The libraries were paired-end (200 bp x 25 bp) sequenced on a HiSeq 2500 instrument (Illumina) using on-board clustering and rapid run mode with a HiSeq PE Rapid Cluster Kit v2 (Illumina) and HiSeq Rapid SBS Kit v2, 200 cycles (Illumina). The HiSeq was running HiSeq Control Software v2.2.68 (Illumina) and Real Time analysis v1.18.66.3 (Illumina). The libraries were prepared and loaded on the HiSeq using the standard procedure (HiSeq 2500 System Guide, Document # 15035786 v01; HiSeq and GAIIx Systems Denature and Dilute Libraries Guide, Document # 15050107 v02; HiSeq System Custom Primers Guide, Document # 15061846 v01) with the following changes. 3.13 µL library pool was denatured by adding 3.13 µL 0.1 N NaOH solution, mixing well by pipetting, and incubating for 5 min at 25°C. The denatured library pool was diluted by adding 993.75 µL cold Hybridization Buffer (Illumina). 400 µL denatured and diluted library pool was mixed with 20 µL denatured and diluted 10 pM Phix Control v3 library (Illumina), and stored on ice until loading. Custom read2 primer mix was prepared by mixing 25 µL SSU_rRNA_read2_fw (100 µM) and 25 µL SSU_rRNA_read2_rv (100 µM) in a conical tube (15 mL) and diluted with 4,950 µL Hybridization Buffer (final concentration is 0.5 µM). When the paired-end reagent rack was loaded on the HiSeq, the Illumina primer mix in position nr. 16 was replaced with the Custom read2 primer mix prepared above. When setting up the HiSeq run in the control software, the standard procedure was followed except for the following steps: For the “Recipe Screen”, the following options were chosen: Index type options = No Index. Read 1 cycles = 200. Read 2 cycles = 25. After sequencing, bcl2fastq v2.17.1.14 (Illumina) was used to generate FASTQ files from bcl files using standard settings (bcl2fastq Conversion User Guide, Part # 15038058 RevB).

Nanopore library preparation and sequencing

Library preparation. Mock community cDNA amplicon product from PCR amplification 2, prepared from 5,000 molecules, was used as input. Nanopore sequencing was carried out following the “Lab protocol for Amplicon sequencing for the MinION device for SQK-NSK007” (Oxford Nanopore Technologies, UK). cDNA amplicon product (340 ng DNA) was diluted to a volume of 45 µL using nuclease-free water. The DNA was transferred to a new tube, and 7 µL Ultra II END-Prep buffer (New England Biolabs, USA) was added, followed by 3 µL End-Prep enzyme mix (New England Biolabs, USA) and 5 µL DNA CS standard. The end repair was incubated at 5 min at 20°C, followed by 5 min at 65°C. The end-repaired DNA was purified using 60 µL Agencourt Ampure XP Bead (Beckman Coulter, USA), and eluting in 31 µL nuclease free water. Adapters were ligated to the end prepared DNA by addition of 8 µL water, 10 µL Adapter Mix, 2 µL HPA, 50 µL NEB Blunt/TA Master Mix (New England Biolabs, USA), and incubated at RT for 10 min. Then 1 µL HPT was added and incubated for 10 min. 50 µL MyOne C1 beads (Thermofisher, USA) were pelleted on a magnet and washed 2x with 100 µL BBB, and the washed beads were resuspended in 100 µL BBB. Ligation cleanup was performed by adding 100 µL washed MyOne C1 beads (Thermofisher, USA) to the adapted, tethered DNA, and incubated at RT for 5 min. Then the beads were pelleted on a magnet and washed 2x with 150 µL BBB. Residual BBB was removed by spinning down any liquid and pelleting on a magnet. The adapted library was resuspended in 25 µL ELB and incubated at 37°C for 10 min. The library size distribution was checked with a Tapestation 2200, using D5000 ScreenTape (Agilent, USA).

MinION Sequencing. A R9 flow cell (Oxford Nanopore Technologies, UK) was mounted in the MinION Mk1B (Oxford Nanopore Technologies, UK), and platform QC was carried out in MinKNOW, revealing 1354 pores. Flow cell priming buffer was prepared by mixing 500 µL RBF1 and 500 µL water. The flow cell was primed twice with 500 µL priming buffer with 10 min in between loadings. The loading mix was prepared by mixing 75 µL RBF1 with 63 µL water and 12 µL Pre-sequencing mix. The sequencing was started in MinKnow (v. 1.0.2) and Basecalling was started using Metrichor "2D Basecalling RNN for SQK-NSK007" (v. 1.107).

Metatranscriptomics

The metatranscriptome was prepared using TruSeq Stranded mRNA Library Prep Kit (Illumina) using standard protocol (TruSeq® Stranded mRNA Sample Preparation Guide, Part # 15031047 Rev. E) with the following changes: A total of 1.5 ng total RNA from activated sludge (A1) was used. The ribo-depletion part of the protocol was skipped, and instead the protocol was started from MAKE RFP step 12 on page 20. Instead of elution from beads, 9.5 µL total RNA sample was simply mixed with 9.5 µL of Fragment, Prime, Finish Mix. The Fragment, Prime, Finish reaction was incubated at 94°C for 12 min. From there on the standard protocol was followed. The resulting metatranscriptomic library was validated using Tapestation 2200 gel electrophoresis and D1000 screentapes. Library concentration was measured using Qubit dsDNA HS Assay Kit on a Qubit Fluorometer 2.0. The metatranscriptomic library was diluted to 4 nM and single-end (240 bp) sequenced on a MiSeq instrument (Illumina) using a MiSeq® Reagent Kit v3, 150 cycles (Illumina). The library was prepared and loaded on the MiSeq using the standard procedure with the following changes. To increase the number of sequencing cycles from 150 to 240 cycles, 3.5 mL incorporation mix was added to well no. 1, and 3.6 mL of scan mix to well no. 2. After sequencing, bcl2fastq v2.17.1.14 (Illumina) was used to generate FASTQ files from bcl files using standard settings.

Metagenomics

The metagenome sample was prepared from DNA extracted from an aliquot of the activated sludge sample (A1) using TruSeq DNA PCR-Free Library Preparation Kit (Illumina). For details on DNA extraction and library preparation see Albertsen *et al*., (2015).

Data availability

Raw and assembled sequencing data is available under the project number PRJXXX. All data processing scripts, documentation and analysis are available on github (Currently only available to reviewers).

Sequencing data processing

Extract “linked-tags” and associated “read-tag” bins. Linked-tags (consisting of 2 read-tags, see **Fig. 1A**) were extracted and processed using a custom perl script (F16S_extract_v4.2.pl). The unique linked-tags were extracted by identifying the conserved flanking adapter positions, counted and sorted by abundance. Any linked-tag with a read-tag present in a more abundant linked-tag variant was discarded. The passing linked-tags were used to associate two individual read-tags and thereby generate linked-tag read bins.

Synthetic long read *de novo* assembly. *De novo* assembly was performed individually on each extracted linked-tag read bin. Parallelized assembly was performed using custom bash scripts and GNU parallel wrapping the following steps. Reads bins were first subsampled to 1000 reads, and trimmed using

cutadapt v1.10[4] with the following command "cutadapt -a AAAAAAAAAAAAA -a TTTTTTTTTTTTT -a GGGCAATATCAGCAC -a GTGCTGATATTGCCC -q 20 -m 50". De novo assembly of each read bin was performed with the overlap assembler SSAKE v3.8.1[5] with the following command "SSAKE -f reads.fa -w 3 -m 20 -o 2 -r 0.51 -p 0 -b output -h -v". The longest contig over 1200 bp from each read-tag bin assembly was concatenated into one file, and used for further analysis.

Trimming and identification of rRNA genes. The assembled sequences were trimmed for residual adapters and to a minimum length of 1200 bp using cutadapt v1.10[4] with the following settings: cutadapt -g TGGTGCTGATATTGC -a GCAATATCAGCACCA -m 1200. The program barrnap (https://github.com/tseemann/barrnap) was used to identify SSU and LSU rRNA genes and furthermore split these to either Bacteria, Archaea or Eukaryotes with the setting --reject 0.3.

Clustering and Chimera checking. Trimmed SSU rRNA sequences were clustered at 97% using USEARCH v. v8.1.1861[6] with the settings "-cluster_fast –id 0.97 –sort size -sizeout" and the centroid of each cluster kept for further analysis. The resulting clustered sequences were chimera checked using USEARCH v. v8.1.1861[6] with the settings "-uchime_denovo" and otherwise default settings.

Nanopore data processing. Fastq files were extracted from the passed 2D reads using Poretools v. 0.5.1[7] and cutadapt v. 1.10[4] used to identify the conserved regions flanking the unique tags in both end of the reads, allowing up to 10% error-rate (equivalent to 1 mismatch). The two unique read-tags from each molecule was extracted and concatenated into a single linked-tag. Afterwards all extracted linked-tags were clustered using USEARCH v8.1.1861[6] with the settings: -cluster_fast –id 0.9. Raw nanopore reads were binned based on their linked-tag cluster and the first 50 bp and 40 last bp of each read removed as they contain adapter sequence. The trimmed reads in each bin were error-corrected using CANU[8]. Each bin with corrected reads were clustered individually using USEARCH v8.1.1861[6] -cluster_fast –id 0.9 and the consensus sequence for each bin exported using the –cons setting. The raw reads and the consensus sequences were mapped to the reference sequences using USEARCH v8.1.1861[6] to estimate sequence identities with the settings: "-usearch_global -id 0.9 -strand plus -maxaccepts 10 -maxhits 1 –userout out.txt -userfields query+target+id+mism+diffs+ql".

Method validation: Mock

In order to estimate the error-rate of the method residual adapters that were missed initially by cutadapt were removed by alignment of the 16S rRNA genes to the SILVA reference alignment using the align.seqs command in Mothur v 1.37.6[9] with default settings. After alignment the sequences were trimmed to span the V1-V9 region of the 16S rRNA gene using screen.seqs with the settings: "start=1046, end=43116". The trimmed 16S rRNA sequences were searched against the 16S rRNA genes of *Escherichia coli* MG 1655 (DSM 18039, NC_000913), *Bacillus subtilis* subsp. *subtilis* str. 168 (DSM 402, NC_000964) and *P. aeruginosa* PAO1 (NC_002516) using USEARCH v8.1.1861[6] with the settings: "-usearch_global -id 0.9 -strand plus -maxaccepts 10 -maxhits 1 –userout out.txt -userfields query+target+id+mism+diffs+ql".

Method validation: PCR primer screening

The full-length SSU rRNA data was screened with commonly used bacteria and eukaryotic primers, to determine the percent of new SSU rRNA sequences that were likely to have been missed in a PCR based effort to obtain SSU rRNA sequences.

Bacteria SSU rRNA sequences clustered at 97% identity were aligned in ARB (see Phylogentic Analysis section), and the alignment manually filtered to extract sequences containing equivalent to E. coli positions 8 to 1492 or more. The primer pair S-D-Bact-0008-a-S-16/ S-D-Bact-1492-a-A-16[10] was tested for hits against the filtered bacteria SSU rRNA sequences, with 0-2 mismatches. The test was performed by using cutadapt v1.10[4] with the following command "cutadapt -g AGAGTTTGATCMTGGC -e Err -O 13 in.fa --discard-untrimmed | cutadapt -a AAGTCGTAACAAGGTA -e Err -O 13 - --discard-untrimmed > out.fa", where Err was equivalent 0-2 mismatches.

Eukaryote SSU rRNA sequences clustered at 97% identity was length-filtered to remove all sequences shorter than 1,350 bp, using cutadapt v1.10[4] with the following command "cutadapt -m 1350 in.fa > out.fa". The primer pair F-566/ R-1200[11] was tested for hits against the filtered eukaryote SSU rRNA sequences, with 0-2 primer/template mismatches. The test was performed by using cutadapt v1.1[4] with the following command "cutadapt -g CAGCAGCCGCGGTAATTCC -e Err -O 13 in.fa --discard-untrimmed | cutadapt -a GCTTAATTTGACTCAACACGGG -e Err -O 13 - --discard-untrimmed > out.fa", where Err was equivalent to 0-2 mismatches.

Method validation: Taxonomic bias

The A1 metatranscriptome and A1 metagenome sequencing data were adaptor-trimmed using cutadapt v1.10[4] with the following command "cutadapt -a GATCGGAAGAGCACAC -a GGAAGAGCGTCGTGT -m 100 -O 13 in.fq > out.fq". The A1 full-length SSU rRNA, the A1 metatranscriptome, and the A1 metagenome data were compared to the MiDAS SSU database v 1.20[12], which is a manually curated activated sludge version of SILVA SSU database (Release 119, Ref NR 99, >1,200 bp). The three datasets were mapped to the MiDAS reference sequences with USEARCH v8.1.1861[6] to extract all the SSU rRNA sequences from the datasets. The following usearch command was used "usearch -usearch_global data.fa -db midas.fa -strand both -id 0.90 -otutabout out.txt". The resulting OTU table was imported into R[13] using RStudio IDE (http://www.rstudio.com/), and the data was filtered to remove eukaryotes and uncultured/unknown families.

Phylogentic analysis

Phylogenetic analyses parameters were based on those applied to construct the recent comprehensive 'Tree of life'[14]. All SSU rRNA sequence datasets were clustered at 97% sequence similarity. An alignment was generated for all sequences >1,200 bp and the 1871 reference sequences from the recent tree of life project[14] using the online SINA aligner[15,16]. Sequences were screened using the alignment quality threshold values applied by SILVA[16] (alignment identity >70%, alignment score >50, base pair score >30). Sequences below threshold were manually assessed against the NCBI sequence database with blastn and excluded if they did not share >70% similarity and >90% coverage with any database sequence. For phylogenetic analysis, a custom ARB filter[17] was applied (filter by base frequency, >5% consensus, gaps counted but not used as maximum) giving 1752 alignment positions. A maximum-likelihood phylogenetic tree was calculated using RAxML v. 8.2.8 with the GTRCAT evolution model[18] on the CIPRES super computer[19]. Affiliation of sequences to major lineages (**Fig. x**) is based on their phylogenetic relationship to classified reference sequences[14] in the constructed tree and their insertion into the SILVA

taxonomic database v1.23.1 NR99[16] with ARB software package[17]. Sequences classified as mitochondrial 16S rRNA genes with the SILVA database were excluded from analyses.

Comparison with the SILVA database

All SSU rRNA sequences were compared to the SILVA database v.1.23.1 NR99[16] using USEARCH v8.1.1861[6] with the following settings: usearch_global –strand both -id 0.6 -query_cov 0.5 -maxaccepts 10 –top_hit_only –userout out.txt -userfields query+target+id+ql+tl+alnlen –notrunclabels".

References

1. Jacobs, M. A. *et al.* Comprehensive transposon mutant library of Pseudomonas aeruginosa. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 14339–14344 (2003).

2. Albertsen, M., Karst, S. M., Ziegler, A. S., Kirkegaard, R. H. & Nielsen, P. H. Back to Basics - The Influence of DNA Extraction and Primer Choice on Phylogenetic Analysis of Activated Sludge Communities. *PLoS One* **10,** e0132783 (2015).

3. Tessier, D. C., Brousseau, R. & Vernet, T. Ligation of single-stranded oligodeoxyribonucleotides by T4 RNA ligase. *Anal. Biochem.* **158,** 171–178 (1986).

4. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17,** 10 (2011).

5. Warren, R. L., Sutton, G. G., Jones, S. J. M. & Holt, R. A. Assembling millions of short DNA sequences using SSAKE. *Bioinformatics* **23,** 500–501 (2007).

6. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26,** 2460–1 (2010).

7. Loman, N. J. & Quinlan, A. R. Poretools: A toolkit for analyzing nanopore sequence data. *Bioinformatics* **30,** 3399–3401 (2014).

8. Berlin, K. *et al.* Assembling large genomes with single-molecule sequencing and locality-sensitive hashing. *Nat. Biotechnol.* **33,** 623–630 (2015).

9. Schloss, P. D. *et al.* Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75,** 7537–41 (2009).

10. Klindworth, A. *et al.* Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* **41,** 1–11 (2013).

11. Hadziavdic, K. *et al.* Characterization of the 18s rRNA gene for designing universal eukaryote specific primers. *PLoS One* **9,** (2014).

12. McIlroy, S. J. *et al.* MiDAS: the field guide to the microbes of activated sludge. *Database (Oxford).* **2015,** bav062– (2015).

13. R Core Team. R: A language and environment for statistical computing. (2016).

14. Hug, L. A. *et al.* A new view of the tree of life. *Nat. Microbiol.* **1,** 16048 (2016).

15. Pruesse, E., Peplies, J. & Glöckner, F. O. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28,** 1823–9 (2012).

16. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41,** D590–6 (2013).

17. Ludwig, W. *et al.* ARB: a software environment for sequence data. *Nucleic Acids Res.* **32,** 1363–71 (2004).

18. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30,** 1312–1313 (2014).

19. Miller, M. A., Pfeiffer, W. & Schwartz, T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. in *Gateway Computing Environments Workshop (GCE), 2010* 1–8 (2010). doi:10.1109/GCE.2010.5676129